Atelier Data Science

Deep learning practice 7

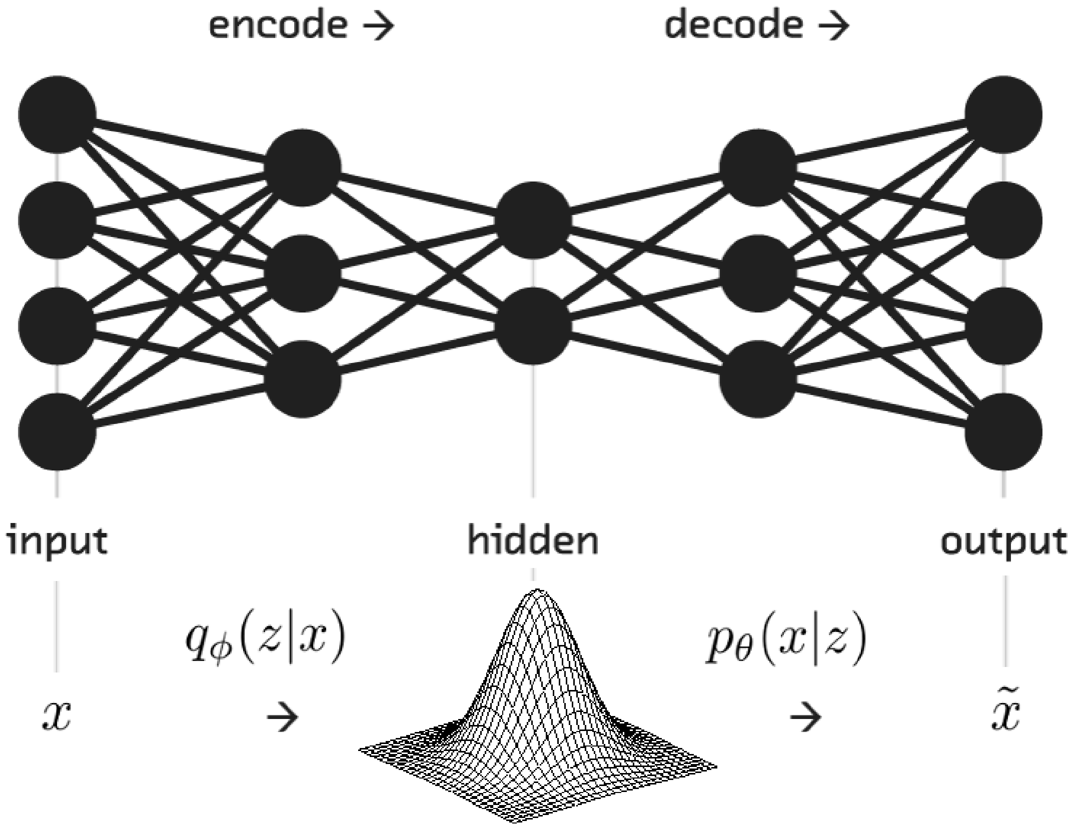
Diffusion Models

Irina Proskurina

Irina.Proskurina@univ-lyon2.fr

Laboratoire ERIC – Université Lyon 2

Variational Autoencoder

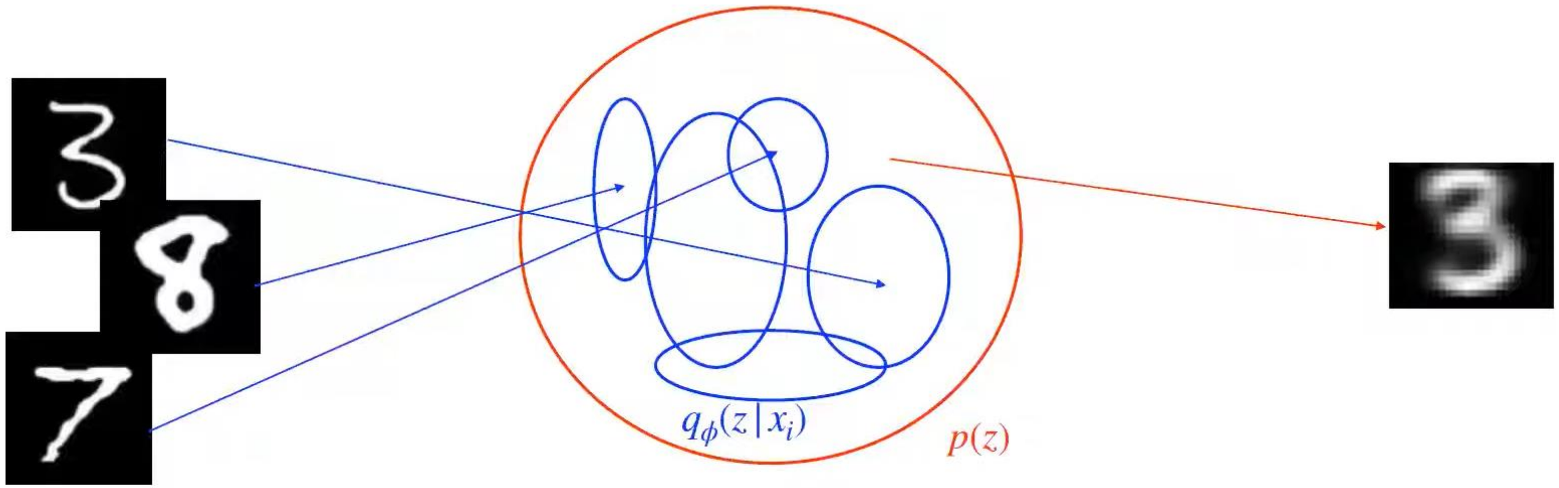


$$\mathcal{L}(\phi, \theta; x) = \mathbb{E}_{z \sim q_\phi(z|x)}[log(p_\theta(x|z))] - \mathcal{D}_{KL}(q_\phi(z|x)||p_\theta(z))$$

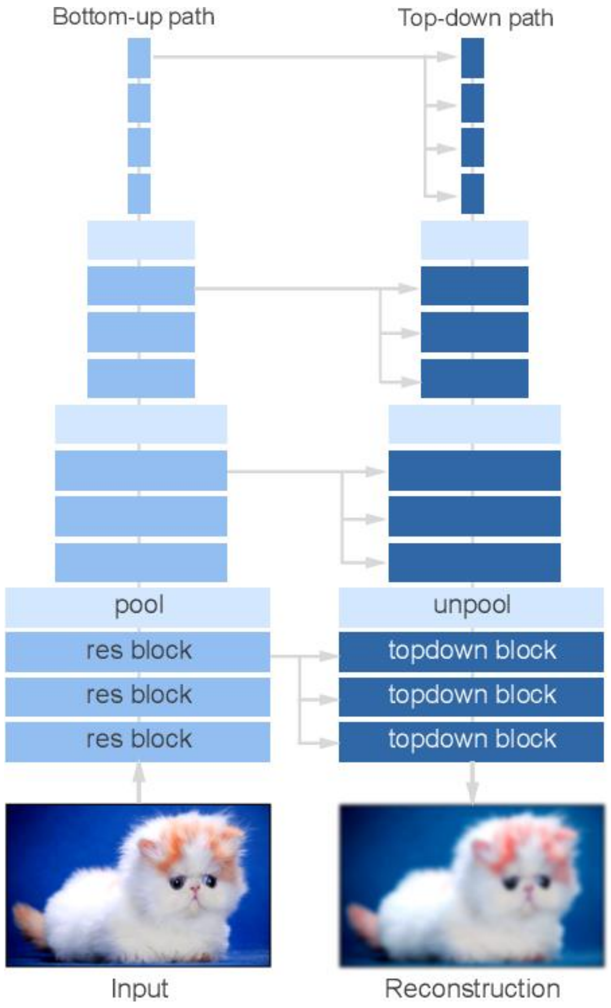
Variational Autoencoder

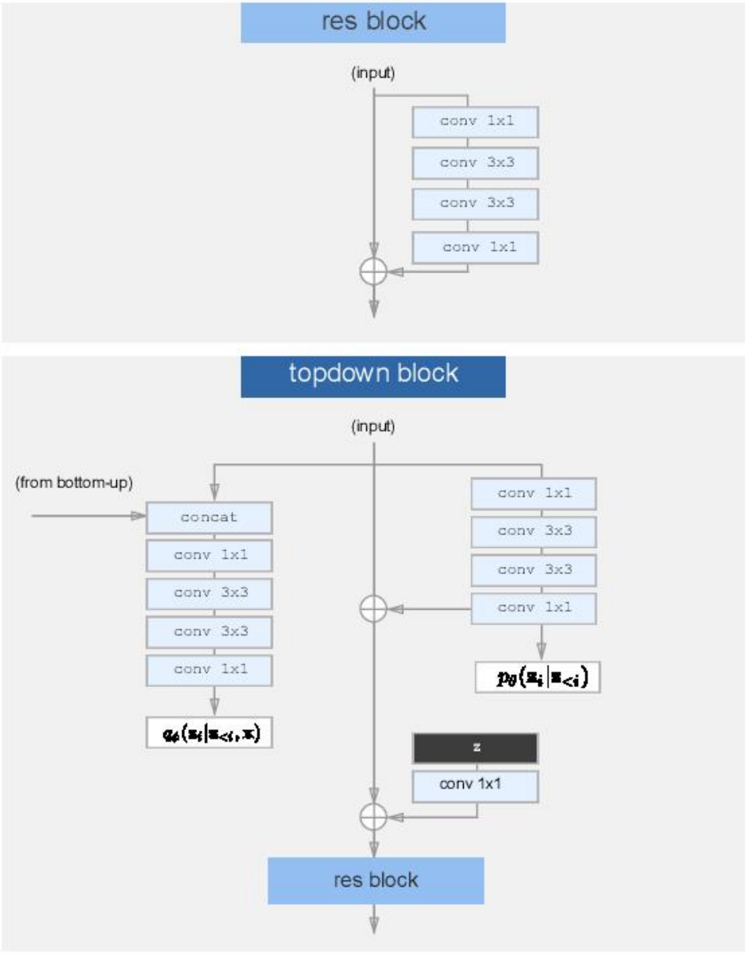
- In VAE we can ensure in good reconstructions but samples from prior may be atypical

$$\mathcal{L}(\phi, \theta; x) = \mathbb{E}_{z \sim q_\phi(z|x)}[log(p_\theta(x|z))] - \mathcal{D}_{KL}(q_\phi(z|x)||p_\theta(z))$$



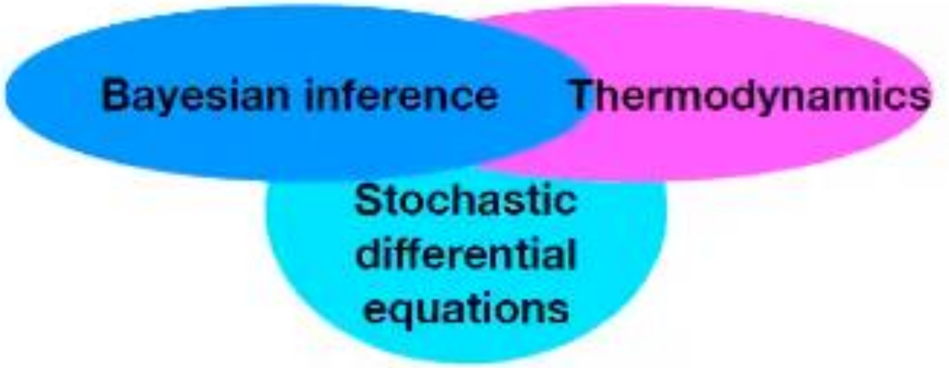
Very Deep VAE





- Very Deep VAE achieved SOTA on image synthesis
- The dimension of latents is much larger than the dimension of observed variables

Diffusion Models



- A breakthrough in generative modeling
- The object is gradually noisified until white noise is obtained
- Reverse process allows to generate objects from noise
- Each object from training set is mapped to exactly the same distribution

Stochastic differential equation

- Ordinary differential equation defines the evolution of a point

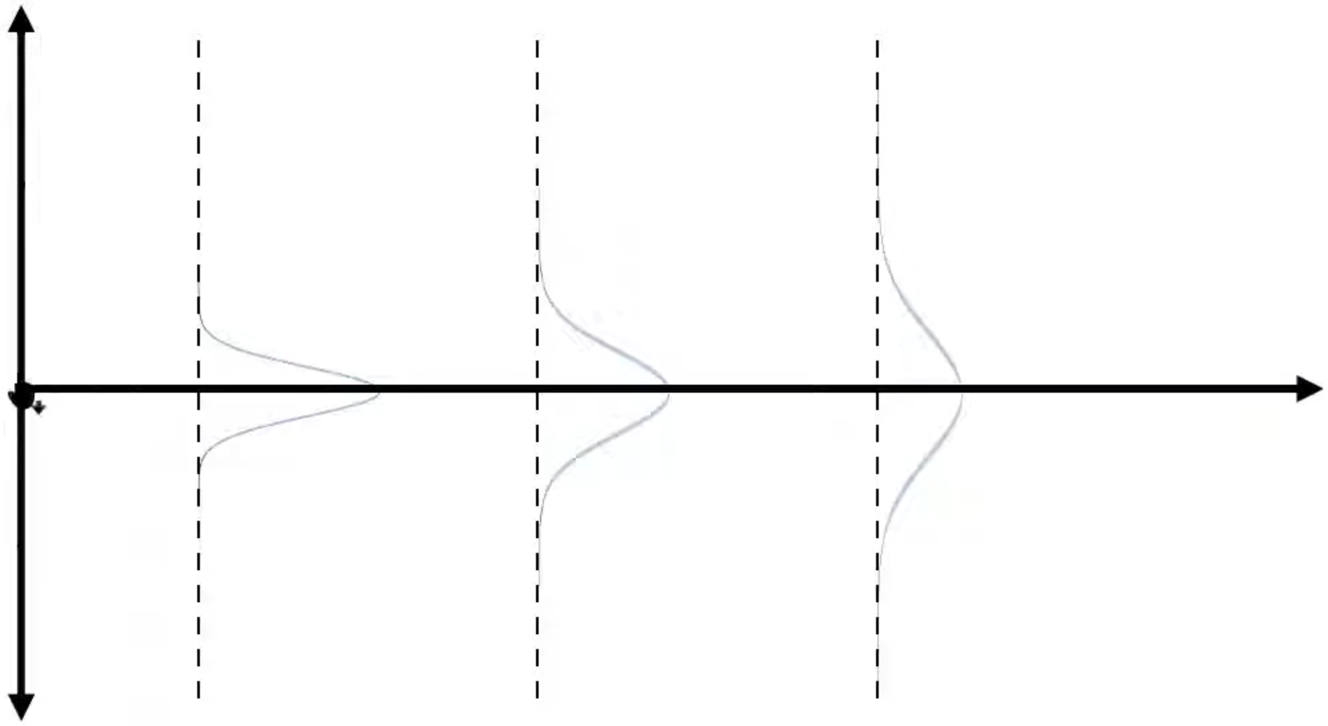
$$dx = f(x, t)dt$$

- Stochastic differential equation defines the evolution of a distribution

$$dx = f(x, t)dt + g(x, t)dW$$

Wiener Process

- The simplest SDE $dx=dW$, $x(0)=0$



Langevin Dynamics

- Langevin Dynamics allows to mitigate the diffusion process
- It uses the gradient of log-density a.k.a. as score function

$$dx = \frac{1}{2}\frac{\partial}{\partial x}\log p_t(x)dt + dW, \quad x(0) \sim p_0(x)$$

- It can be shown that under such dynamics the distribution does not change

$$\frac{\partial}{\partial t}p_t(x) = 0$$

Backward dynamic

- Consider SDE that maps some complex distribution to white noise

$$dx = \alpha x dt + dW, \quad \alpha < 0, \quad x_0 \sim p_0(x)$$

- Let $p_T(x)$ be a distribution at time *t*
- Is it possible to reconstruct initial distribution from noise? Yes! By using backward dynamics SDE (Song20)

$$dx = [\alpha x - s(x,t)]dt + dW, \quad \alpha < 0, \quad x_T \sim p_T(x)$$

- Note that backward SDE essentially uses the score

$$s(x,t) = \frac{\partial}{\partial x} \log p_t(x)$$

Deterministic dynamics

- It can be shown that the equivalent dynamics is set by deterministic ODE

$$dx = \left(\alpha x - \frac{1}{2}s(x,t)\right)dt$$

- Recall (forward) Langevin dynamics $dx = \frac{1}{2}s(x,t)dt + dW$

- If we add forward Langevin we obtain forward diffusion process $dx = \alpha x dt + dW$

- If we add backward Langevin we obtain backward dynamics SDE

$$dx = (\alpha x - s(x,t))dt + dW$$

Training process

- We may consider (discretized) diffusion model as a special case of hierarchical VAE

- Let x0 be observed variable and X1, ..., Xt are latents. Then

$$q(x_1, \dots, x_T | x_0) = q(x_T | x_0) \prod_{t=1}^{T-1} q(x_t | x_{t+1}, x_0), \quad p_\theta(x_0, \dots, x_T) = p(x_T) \prod_{t=0}^{T-1} p_\theta(x_t | x_{t+1})$$

$$\mathscr{L}(\theta) = \mathbb{E}_{x_1, \dots, x_T | x_0} \log p_\theta(x_0 | x_1) - \sum_{t=1}^{T-1} \underbrace{\boxed{\mathbb{E}_{x_{t+1}, \dots, x_T | x_0} KL(q(x_t | x_{t+1}, x_0) \,||\, p_\theta(x_t | x_{t+1}))}}_{\text{Score matching term}}$$

Generative Models